Homework3 Report — Image Recognition

學號：R06323011 系級： 經濟碩一 姓名：葉政維

1. **(1%) 請說明你實作的 CNN model，其模型架構、訓練參數和準確率為何？(Collaborators: 經濟碩二田家駿，感謝提供我一些通過 baseline 的大方向建議以及總訓練參數數的想法)**

實作的 CNN 模型架構為:conv(36,3,3)*2 — maxpool(2,2) – conv(72,3,3)*3 — maxpool(2,2) — flatten() — dropout(0.25) — dense(400) — dense(40) — dense(7)。Activating function 除輸出層外，皆為 relu。此外在每層 conv2d 之後皆進行 batch normalization。總共可訓練參數為 1989419 個，而 private accuracy 為 65.895。

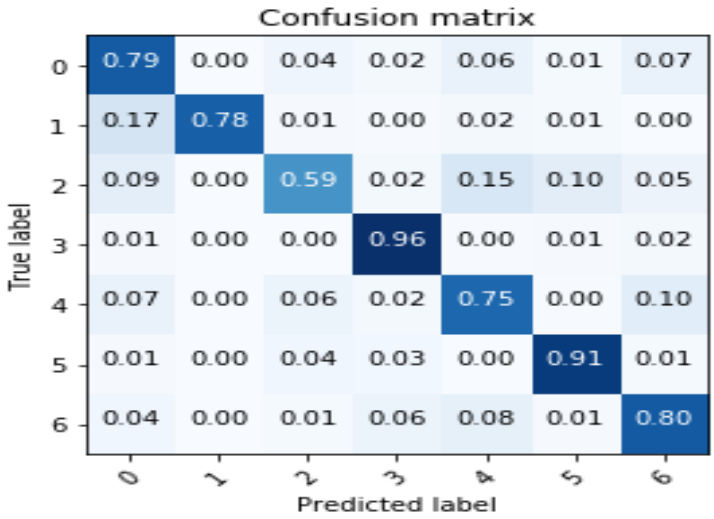
架構的設計是參考自 VGG 以及網路上的一些範本，疊加 conv 以及倍數成長的 filter 似乎是其中的特色。為了不讓本人的筆電花費過多時間，因此架構設計、epoch 等有經過一些調整與簡化。

2. (1%) 請嘗試 data normalization, data augmentation,說明實行方法並且說明對準確率有什麼樣的影響？

為方便訓練並進行比較，本題使用較為簡化的模型。為實作 data normalization，先將整個 dataset，逐 pixel 減去平均並除上標準差；而為實作 data augment，則是利用 ImageDataGenerator(…)對樣本進行若干轉變，如水平翻轉、左右平移等。

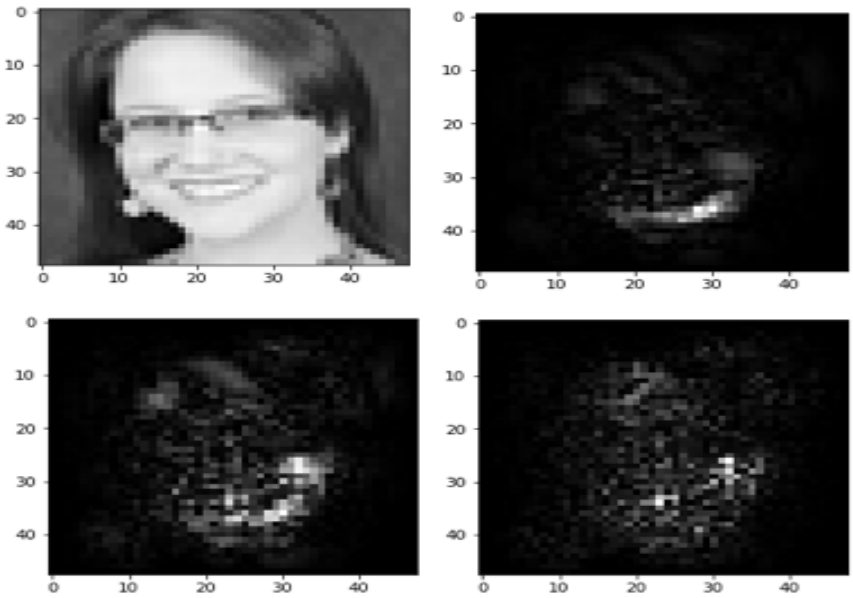
完全沒有處理過、額外 normalize 以及額外 data augment，其 private 準確率分別為 56.976、57.676、61.827。大致上 data augment 可以讓模型表現提升許多，原因如課堂上提到的：不要讓模型對於一些無關緊要的特徵有反應。Normalization 只有些微的幫助，不過似乎比未 normalize 的做法更早收斂到最佳解，這點有如課堂上提到 Normalization 可以讓 gradient descend 的過程更平滑。

3. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？[繪出 confusion matrix 分析]



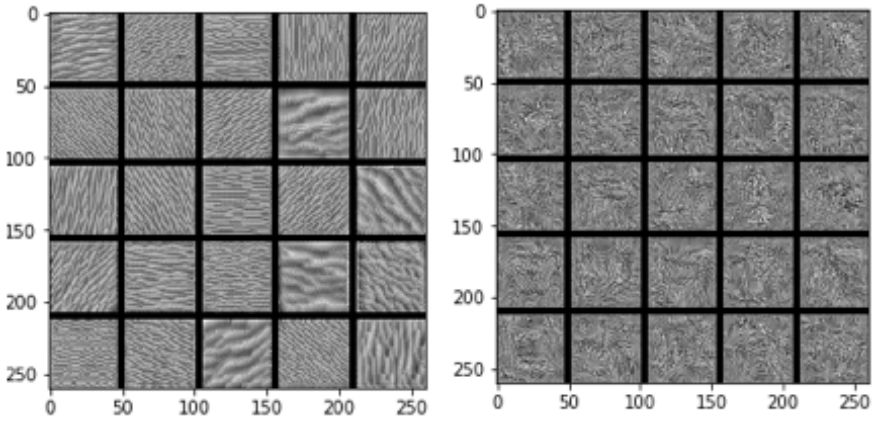
上圖使用 sklearn confusion matrix 的參考文件繪製，並沿著 true label 做 normalize。經觀察，fearful(2)是較難分類的一類。事實上，經本人檢閱原始圖片，有許多看起來是 worried(4)或 surprised(5)的圖片，事實上是 fearful。不過有些 worried 以及 surprised 的圖片是相對容易辨識的（如那種眼睛瞪得大大的，嘴巴大開），這點能從上圖分錯的比例中得到一些驗證。最後，如同個人觀察，happy(3)是相對好判斷的一群。

4. (1%) 從(1)(2)可以發現，使用 CNN 的確有些好處，試繪出其 saliency maps，觀察模型在做 classification 時，是 focus 在圖片的哪些部份？

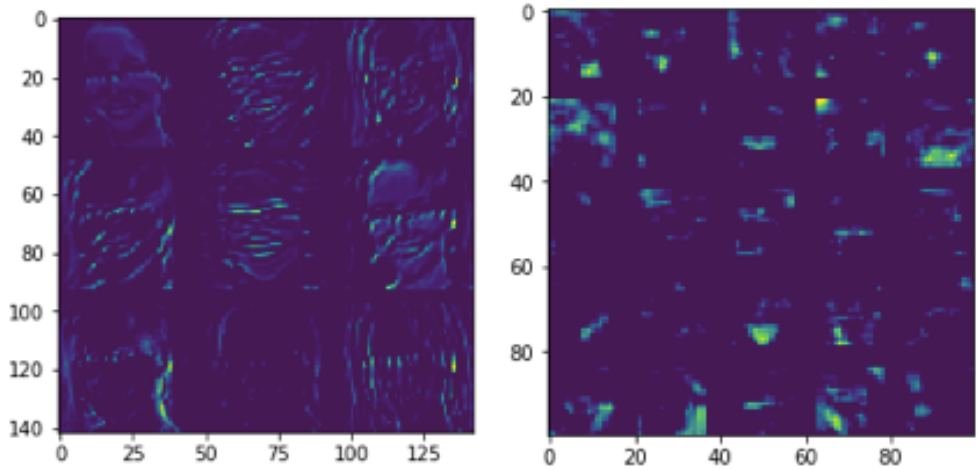


上圖為取用其中一張標籤為 happy 的圖，繪製 saliency map。繪製方式是針對每個 pixel 進行些微變動，觀察哪些變通會造成輸出層 (happy) 的變動最多。右上至右下分別是使用不同的”微變動”進行繪製。整體而言，這些圖大概在嘴部以及因為微笑而隆起的臉頰特別有反應，和我們對於笑臉（開心）的觀察十分符合。

5. (1%) 承(1)(2)，利用上課所提到的 gradient ascent 方法，觀察特定層的 filter 最容易被哪種圖片 activate。



參考作業說明連結提供的 code，繪製各 filter 所對應，最能 activate 該 filter 所對應 output 的 input。上圖左、右是分別使用第二、第五層 conv2d（見題一），取前 25 名活躍的 filters。上圖左大多是一些條紋狀並別的圖形，有可能是在擷取一些重要特徵的輪廓，上圖右則較難以辨識，有可能是在擷取一些零碎、拼湊的特徵，這對人類可能較不直觀。



上圖則選用和題 (4) 一樣的圖片，觀察其在上述 layers 的輸出，左上的圖（僅取九筆）可以觀察到人的輪廓，且不同圖間不同走向的輪廓強度也有差異，如有幾張圖橫向的輪廓更為明顯。右上的圖，也許是因為在較後的 layer，不僅解析度較低，似乎也比較專注在局部特徵而非整體概況的擷取。